Analysis of ego-network

Performed by Lebedeva Anna

Data

Data was scraped from VK API

The steps were:

1. to collect list of my friends
2. to collect a dictionary with my friends' friends
3. leave only mutual friends of ours and create a list of pairs, which will be the edges of the graph
4. extract additional information from the API: names and universities

Example of the result:

```
1  gg.nodes[200]
```

```
{'name': 'Андрей Стрельников', 'uni': 'no info'}
```

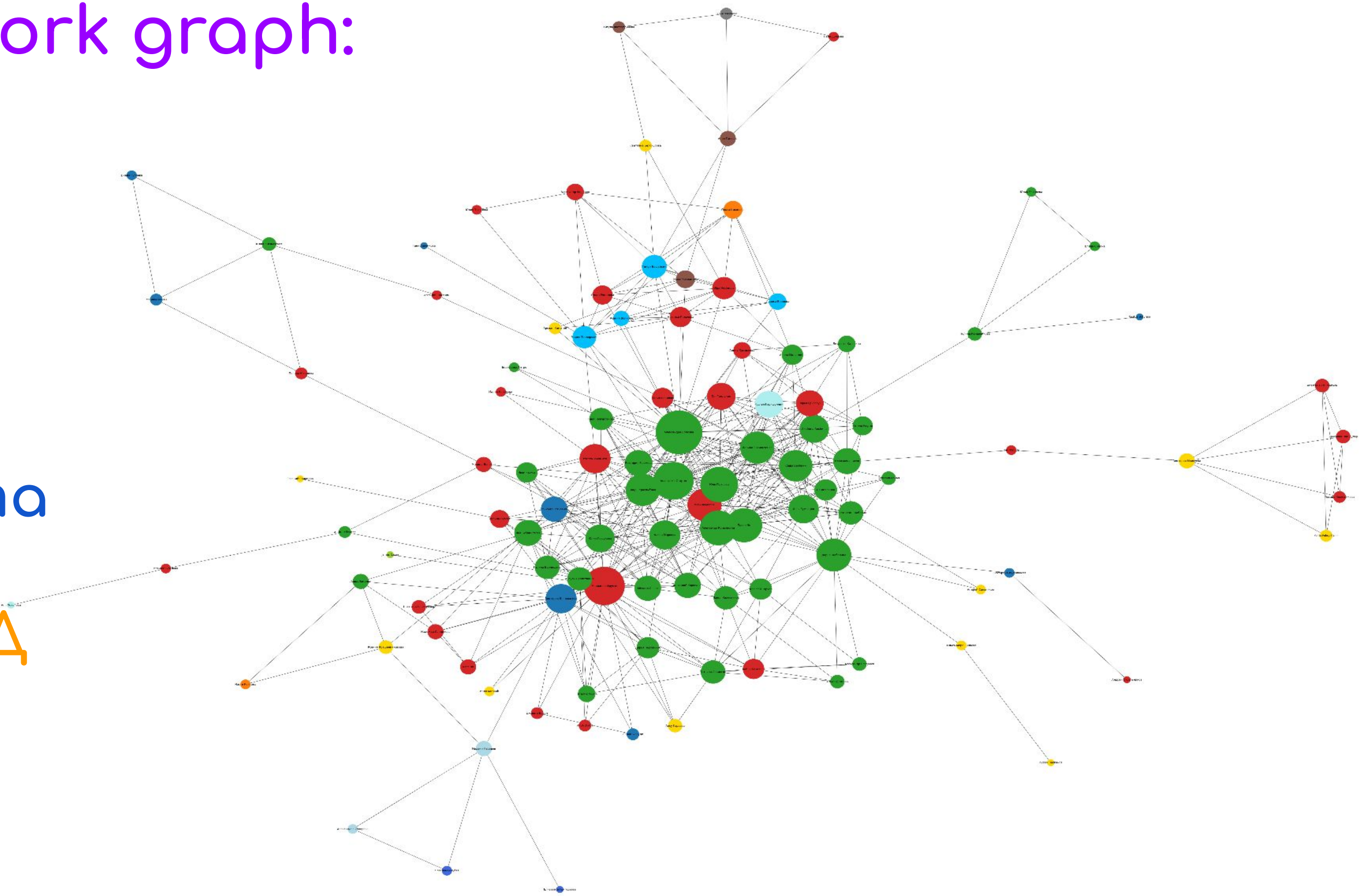
Preparing the graph:

The preparation steps were:

1. to work on the list of Universities, choose the most common ones and assign ‘other’ label to the rest
2. assign colors to Universities
3. assigning these colors and nodes names as attributes to the graph
4. creating degree attribute to set a size for the nodes

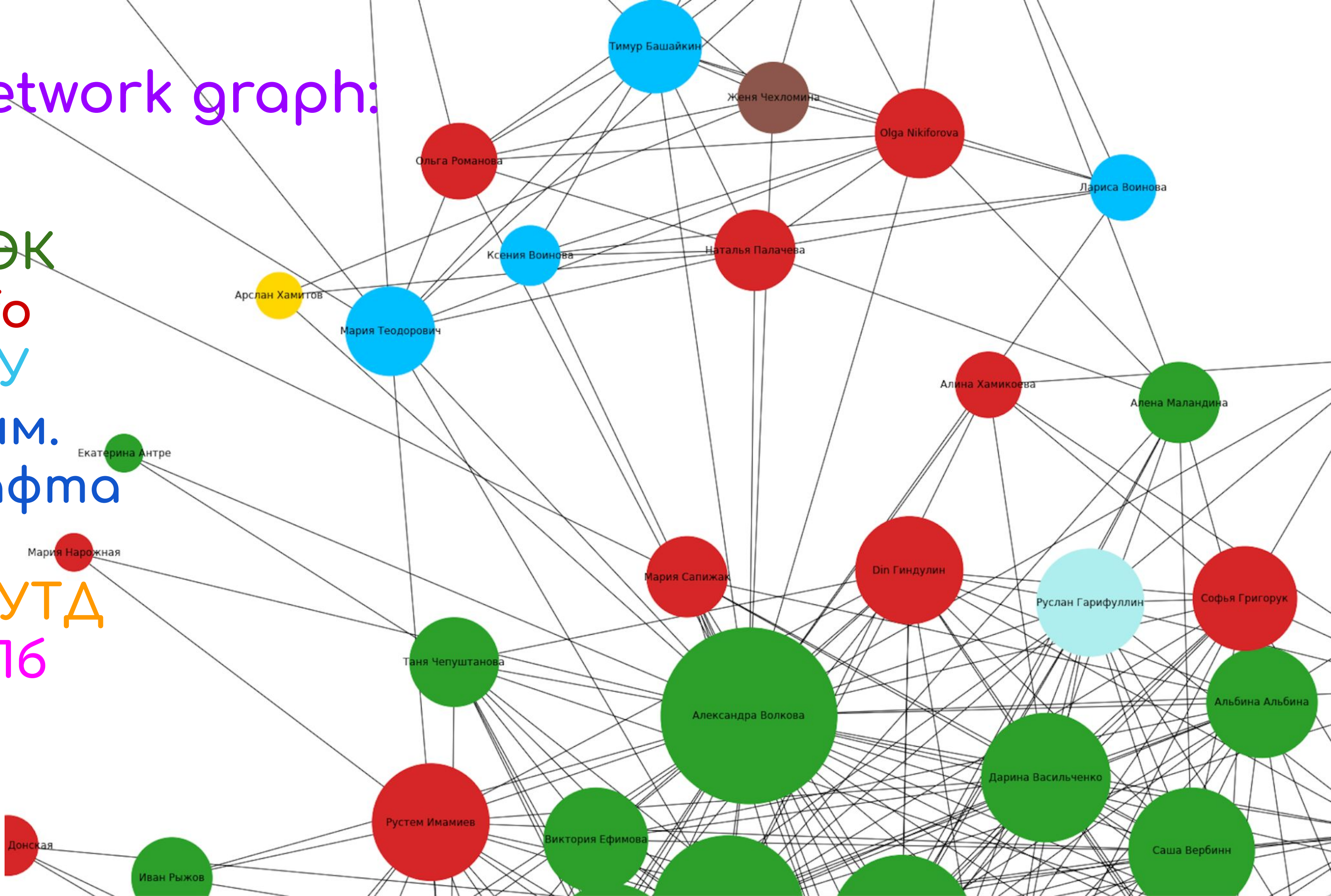
Ego-network graph:

- ФИНЭК
- no info
- СПбГУ
- НГУ им. Лесгафта
- КФУ
- СПбГУТД
- ЕУ СПб
- other



Ego-network graph:

- ФИНЭК
- no info
- СПбГУ
- НГУ им. Лесгафта
- КФУ
- СПбГУТД
- ЕУ СПб
- other



Simple statistics:

Number of nodes: 108

Number of edges: 476

Diameter: 8

Radius: 4

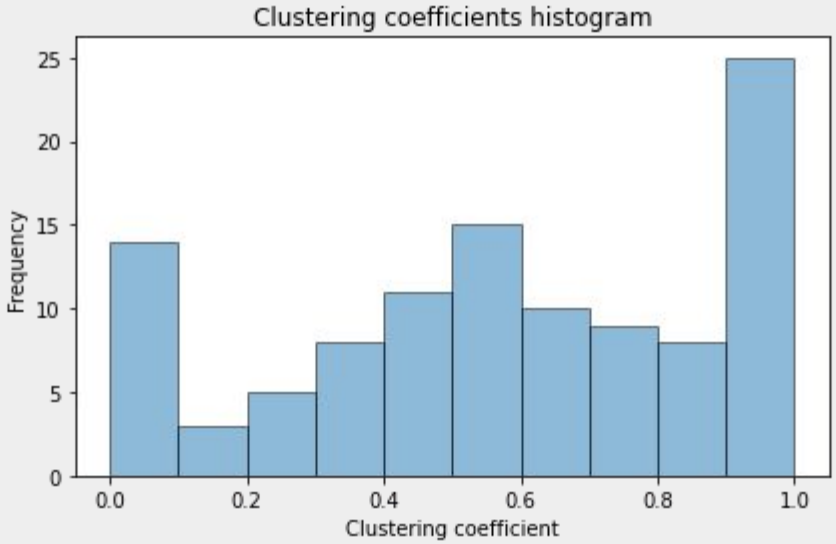
Average shortest path length: 3.1478020076150917

Average clustering coefficient: 0.5773251551486372

Clustering coefficients analysis:

Conclusion:

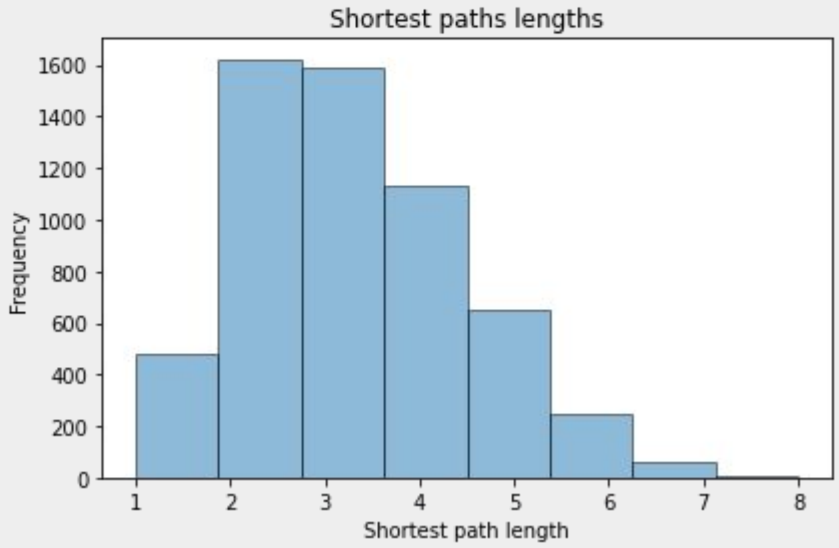
There are nodes that are tightly connected, but some nodes are more separate.



Analysis of shortest path lengths between each pair of nodes:

Conclusion:

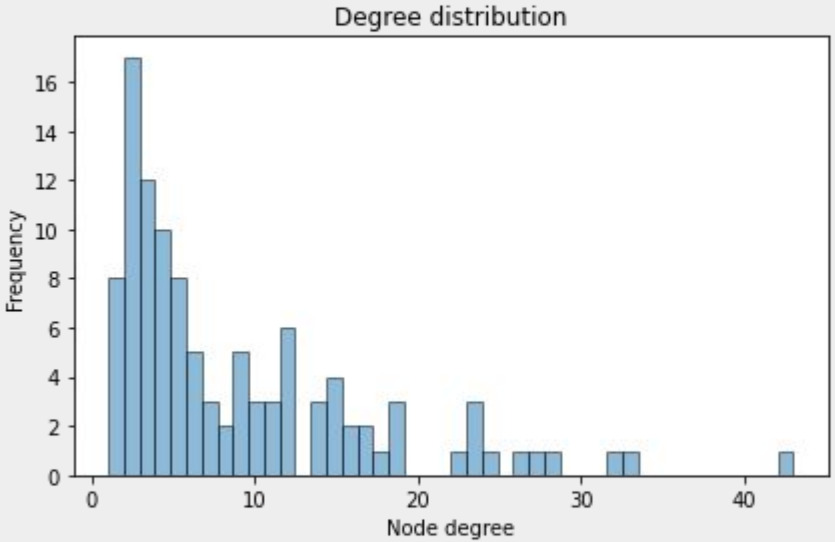
- 8 is a longest “shortest path”, which means the graph is quite tightly connected
- Most paths are in interval of 2-3 edges



Analysis of degree distribution:

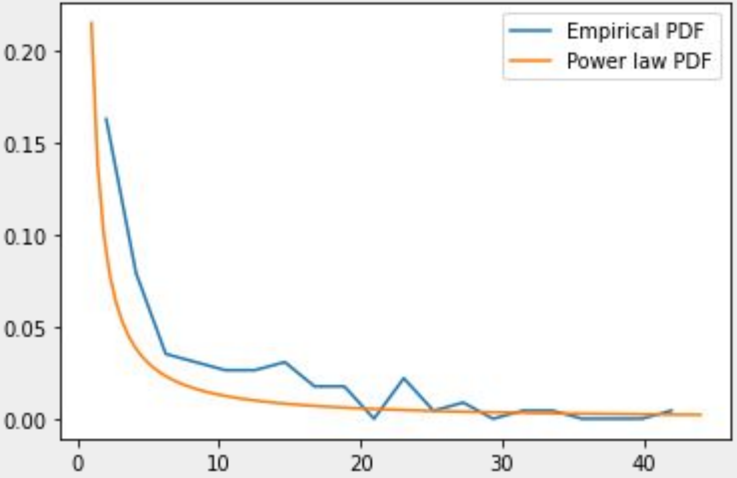
Conclusion:

- Most nodes have degree from 1 to 5
- By the eye it does look similar to Power law distribution



Probability density function (PDF):

- alpha for power law function = 1.2154797693074315
- KS-test p-value = 0.00013402951912666605
- Null hypothesis rejected



Degree/Closeness/Betweenness/Eigenvector centralities:

Degree:

	person	centrality
6	Александра Волкова	0.401869
17	Марианна Фрунзе	0.308411
9	Анастасия Старых	0.299065

The **degree** of a node of a graph is the number of edges that are incident to the node. Which means, the people who are represented by central nodes, counting by the degree, have the biggest number of friend among my friends. So this means we have biggest amount of mutual friends with them. Which is proved by the actual data: with Alexandra we have 43 mutual friends, with Marianna - 33 mutual friends, with Anastasia - 32 mutual friends.

Degree/Closeness/Betweenness/Eigenvector centralities:

Closeness:

	person	centrality
6	Александра Волкова	0.529703
9	Анастасия Старых	0.495370
23	Юля Пытьева	0.461207

The more central a node is, the closer it is to all other nodes. Which means, if we want to spread the information to the whole network, but we need to choose, for example, 3 people, to whom we will tell it, and we consider that information spreads with equal time between each node, these are three people we need to choose to tell the information to. This doesn't mean that we have biggest amount of mutual friends with these people, but it means that they have shortest access to other people in the network.

Degree/Closeness/Betweenness/Eigenvector centralities:

Betweenness:

	person	centrality
6	Александра Волкова	0.332067
9	Анастасия Старых	0.131487
17	Марианна Фрунзе	0.124328

The betweenness centrality for each node is the number of shortest paths that pass through the node. This means that the biggest amount of information goes through these nodes in the network. So maybe lifewise it means that these people know the most information about all other people in the network altogether.

Degree/Closeness/Betweenness/Eigenvector centralities:

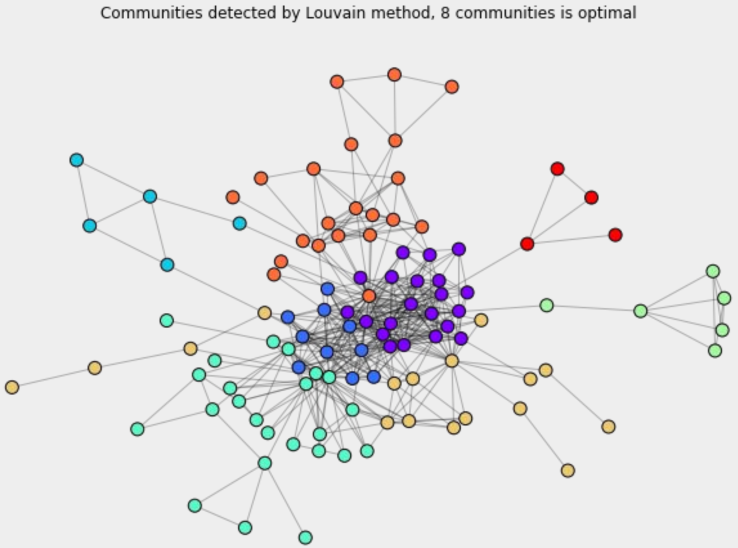
Eigenvector

centrality:

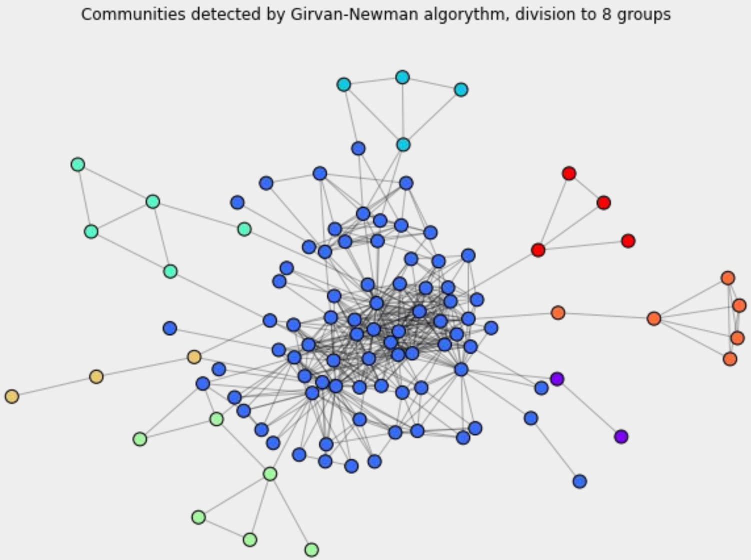
	person	centrality
6	Александра Волкова	0.271048
23	Юля Пытьева	0.254491
9	Анастасия Старых	0.247444

Eigenvector centrality is a measure of the influence of a node in a network. It is based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes. A high eigenvector score means that a node is connected to many nodes who themselves have high scores. I am not sure how to interpret it in terms of ego network, but I may suggest that if we want the network to agree on some point, the most effective way will be to persuade the people who are represented by nodes with biggest eigenvector centralities.

Comparing community detection algorithms:



Louvain method

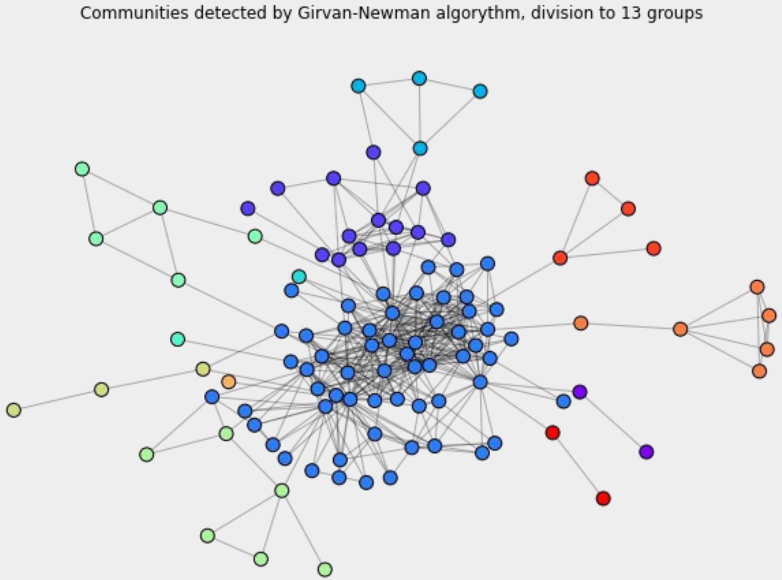


Girvan-Newman algorithm

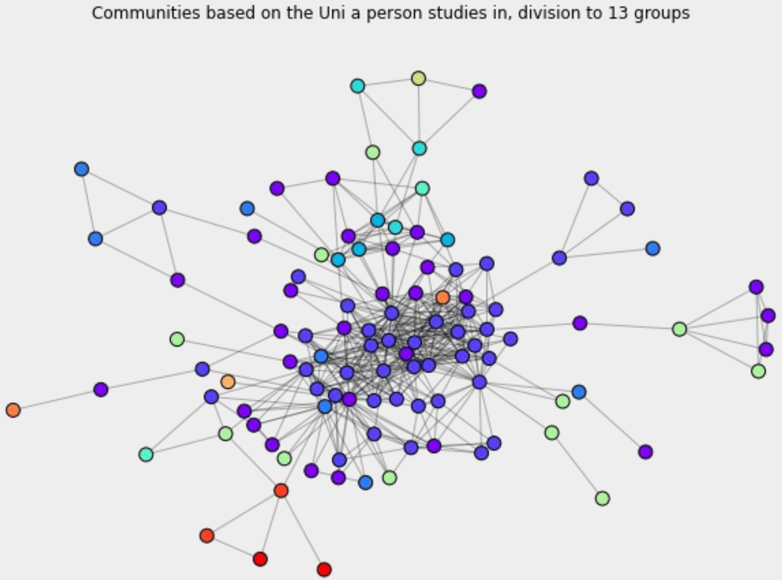
Comparing community detection algorithms:

	Louvain	Girvan-Newman
tend to separate nodes on the sides	✅	✅
is based on:	modularity	edge betweenness
tends to:	drags more nodes from the center	leaves more nodes in the center
Works better, judging by my criteria: Knowledge coefficient	❌	✅

Comparing Girvan-Newman algorithm to the division by Uni:



Girvan-Newman algorithm



Uni division

Comparing Girvan-Newman algorithm to the division by Uni:

Knowledge coefficient of Girvan-Newman clustering: 0.4865795980757991
Knowledge coefficient of Uni clustering: 0.13557636924331623

Conclusion:
We see that the Knowledge coefficient for Uni division is very low, so University a person studied in is not an argument for them to belong to a certain community. Partly these results may be not correct because I combined some Universities into "other" category, and also for many nodes there is no such information.

Thank you for attention!